

### **REMARKS**

The Applicant has now had an opportunity to carefully consider the comments set forth in the Office Action mailed August 10, 2006. All of the rejections are respectfully traversed. Amendment, reexamination and reconsideration of the application in view of the following remarks are respectfully requested.

#### **The Office Action**

In the Office Action mailed August 10, 2006:

**claims 2, 7 and 23** were rejected under 35 U.S.C. §102(b) as being anticipated by U.S. Patent No. 5,963,205 to Sotomayor ("Sotomayor");

**claims 4, 5, 8, 10-13, 18 and 20-28** were rejected under 35 U.S.C. §103(a) as being unpatentable over Sotomayor in view of U.S. Patent No. 6,356,840 to Saito, et al. ("Saito");

**claim 6** was rejected under 35 U.S.C. §103(a) as being unpatentable over Sotomayor in view of U.S. Patent No. 5,276,616 to Kuga, et al. ("Kuga");

**claims 15 and 16** were rejected under 35 U.S.C. §103(a) as being unpatentable over Sotomayor in view of Saito and further in view of U.S. Patent No. 4,903,229 to Schmidt, et al. ("Schmidt"); and

**claim 17** was rejected under 35 U.S.C. §103(a) as being unpatentable over a combination of four references including Sotomayor, Saito, Schmidt and U.S. Patent No. 6,064,397 to Herregods, et al. ("Herregods").

#### **The Present Application**

By way of brief review, the present application is directed to systems and methods for automatic and semi-automatic document indexing. For instance, the invention is useful where a large document is scanned to generate an electronic version of the document. For example, the invention is used to generate a table of contents or an index for the electronic version of the document (for example, see FIG. 1). A custom sub-section delimiter definition that is related to the document currently being processed can be generated. The delimiter definition includes a list of one or more characteristics of a delimiter (page 5, lines 13-30).

For example, a review of the document may indicate that chapter headings in the document are rendered in an 18-point font size at a location that is centered

horizontally on a page and is two inches below the top of the page. In that case, a first sub-section delimiter may be defined as any text located two inches from the top of a page and rendered in an 18-point font size.

Subheadings in the exemplary document might occur anywhere on a page but are rendered in a 16-point font size with underlined characters. Therefore, a second sub-section delimiter for the document might be defined as underlined 16-point text.

Once one or more sub-section delimiters are defined, the electronic version of the document is searched to find occurrences of text corresponding to the defined sub-section delimiters. Information regarding each occurrence is used to create an index or table of contents for the document. For example, for each occurrence of 18-point text located two inches from the top of a page, the text string associated with the occurrence is recorded in association with a text location of the occurrence. For example, the text "CHAPTER ONE" was found rendered in 18-point font two inches from a top edge of page 3 of the document. Therefore, the text "CHAPTER ONE" is associated with a page number, such as page 3, and is recorded and included in the index or table of contents. The text location information may be recorded in the form of a hypertext link. Similarly, the text and text location associated with each occurrence of underlined 16-point text is recorded and added to the index or table of contents. For example, the text location recorded may include a page number as well as an indication of a location within the page. Again, the text location information may be recorded in the form of a hyperlink.

### **The Newly Cited Reference**

In stark contrast, the newly cited and primary reference of the Office Action to Sotomayor allegedly discusses automatic index creation for a **word processor**. Sotomayor allegedly discloses a system and method that automatically identifies key topics and phrases in a document text and inserts identifying tokens for the index generation program **in the word processor** to generate an index to those key topics (column 2, lines 41-45). Allegedly, a one embodiment of the invention of Sotomayor includes a process running **in a word processor program** on a computer which (a) allows an author to select index generation for a document being processed (edited) and then, **using a semantic analyzer program** running on a computer, (b)

automatically identifies significant key topics within the document, (c) generates and embeds index tokens into the text of the document (column 2, lines 54-61). Even if the tokens of Sotomayor are fairly construed as predetermined machine readable symbols, as the phrase is used in the present application (which is disputed), it is respectfully submitted that Sotomayor does not disclose or suggest determining a subsection delimiter comprises a user indicating a predetermined machine readable symbol or a user placing a predetermined machine readable symbol representing a demarcation point on a printed version of the document.

It is respectfully submitted that the secondary references do not cure the deficiencies of Sotomayor.

For example, the newly cited secondary reference to Saito allegedly discusses a user-defined search template for extracting information from documents. However, Saito does not disclose or suggest a user indicating a font size or **a user indicating** a font style or **a user indicating** a text string to be used as a subsection delimiter or **a user indicating** a text location or a user indicating a specific point coordinate within a document or **a user placing** a predetermined machine-readable symbol representing a demarcation point on a printed version of the document.

Even if FIG. 15 of Saito indicates that a search template can include coordinates, indentation, font size and font type, Saito does not disclose or suggest that a user indicates these pieces of information. Instead, in the system of Saito, the user is allowed to select divided areas where desired information is contained and to provide a user-defined name or label for each of the selected areas (column 4, lines 40-44).

In the system of Saito, it is a logic structure construction unit 1107 that extracts associated characteristic information for the selected areas (column 4, lines 44-46). Saito later reiterates that the coordinates of the text area, font information, such as font size and font type of the characters in the text area, and the layout information, such as indentation and columnar arrangement are collected for each text area (column 7, lines 51-55). Accordingly, it is respectfully submitted that Saito does not disclose or suggest a user indicating at least one of (i.e., and perhaps only one of) a font size, a font style, a text string, a text location description, a specific point coordinate within a document or a user placing a predetermined

machine-readable symbol representing a demarcation point on a printed version of the document as recited, for example, in claim 2 of the present application. Accordingly, as explained in reference to FIG. 17 and FIG. 18, a plurality of samples are processed according to the preferred process of Saito so that weight coefficients based upon the reliability of the layout characteristics determined by the logic model generation unit can be determined (column 9, lines 9-40). In this regard, it is respectfully submitted that generating the search template of Saito is far more complicated and time consuming than determining a subsection delimiter as recited, for example, in claims 1, 7, 15, 11 and 23 of the present application.

Additionally, it is respectfully submitted that there is no motivation in the art to combine subject matter from Saito with subject matter from Sotomayor other than that which could be gleaned from the present application. Accordingly, rejections of the claims in the present application are based on impermissible hindsight.

Kuga, Schmidt and Herregods have been addressed in previous papers submitted by the Applicant including, for example, Applicant's Amendment D, which was mailed by the Applicant on August 6, 2004.

#### **The Claims are not Anticipated**

**Claims 2, 7 and 23** were rejected under 35 U.S.C. §102(b) as being anticipated by Sotomayor.

In explaining these rejections, the Office Action draws an analogy between HTML heading tokens, which allow different computers to process HTML documents for visual presentation in a manner customized for particular display devices (column 6, lines 5-15), with the predetermined machine-readable symbols recited in, for example, **claims 2 and 23** of the present application. However, it is respectfully submitted that disclosure of a heading token being included in a document does not disclose or suggest a user indicating that that heading token should be used as a subsection delimiter for searching the document and generating an index for the document.

In the system of Sotomayor, a linguistic analyzer translates the input document into a data structure, such as an Intelligent Paragraph Format that is shown in FIG. 9A. The Intelligent Paragraph Format (IPF) consists of word objects and paragraph objects. Each word object **800** contains a stem entry stem **801** (e.g.,

"open") a suffix **802** (e.g., "-ing"), a syntactic value **803** (e.g., "15" for a singular noun) and a semantic weight **804**. Each word object is generated and obtained from a regular lexicon dictionary **195**. One embodiment of regular lexicon dictionary **195** contains over 110,000 word objects **800**, each word object **800** having its own syntactic value **803**, semantic weight **804** and suffix **802** (column 14, lines 8-18). Accordingly, it is respectfully submitted that disclosure of the IPF paragraph objects for the source document 20 being scanned to find all the headings in a document at column 15, lines 60-63, cited by the Office Action, does not disclose or suggest searching for the heading tokens of Sotomayor. Moreover, Sotomayor does not disclose or suggest the user indicating that the heading tokens should be searched for (i.e., are an element of a subsection delimiter for which corresponding occurrences should be searched).

Column 15, lines 63-65, indicate that the number of headings output is specified by the spinner at 111, and is a number between 1 and 6 (i.e., not by the user). In conjunction with disclosure at column 16, lines 1-5, cited by the Office Action indicates that one embodiment of summary page generator 40 always generates all six heading levels rather than giving the user the ability to select the number of heading levels only serves to support the assertion of the Applicant that the system of Sotomayor does not disclose or suggest **a user indicating** elements of a subsection delimiter definition (i.e., as recited, for example, in **claim 2**).

For at least the foregoing reasons, **claims 2, 7 and 23** are not anticipated by Sotomayor.

In any event, **claims 2 and 23** have been amended to remove their references to a user indicating a predetermined machine-readable symbol, and **claim 7** does not include such a recitation. Additionally, the Office Action does not assert that Sotomayor discloses or suggests, and it is respectfully submitted that Sotomayor does not disclose or suggest, a user indicating at least one of a font size, font style, text string, text location description and a specific point coordinate within the document, or wherein determining a subsection delimiter comprises a user placing a predetermined machine-readable symbol representing a demarcation point on a printed version of the document as the subsection delimiter, as recited, in **claim 2** of the present application. Additionally, it is respectfully submitted that Sotomayor does not disclose a user selecting an exemplary subsection title,

performing one of document recognition and optical character recognition on the selected exemplary subsection title to determine at least one recognized property and using the at least one recognized property of the exemplary subsection title as a subsection delimiter, as recited, for example, in **claim 7**, or a user indicating at least one of a font size, a font style and a specific point coordinate within the document as recited in **claim 23**. Accordingly, it is respectfully submitted that **claims 1, 7 and 23** are not anticipated by Sotomayor.

For at least the foregoing reasons, it is respectfully submitted that **claim 2**, as well as **claims 4-6**, which depend therefrom; **claim 7**, as well as **claim 8**, which depends therefrom; and **claim 23**, as well as **claim 25**, which depends therefrom, are not anticipated and are not obvious in light of Sotomayor.

#### **The Claims are not Obvious**

**Claims 4, 5, 8, 10-13, 18 and 20-28** were rejected under 35 U.S.C. §103(a) as being unpatentable over Sotomayor in view of Saito.

In explaining these rejections, the Office Action reiterates the assertions made with regard to **claims 2, 7 and 23**. In this regard, arguments similar to those submitted in support of **claims 2, 7 and 23** are submitted in support of **claims 4, 5, 8, 10-13, 18 and 20-28**. Disclosure of an author including heading tokens in a document so that browsers may process the document for visual presentation in a manner customized for a particular display device (column 6, lines 10-12) does not disclose or suggest a user indicating that one or more of the tokens should be included as part of a subsection delimiter definition for which occurrences corresponding thereto should be searched for in order to generate an index. Furthermore, Sotomayor does not disclose or suggest that the heading tokens of Sotomayor are IPF paragraph objects for which the system of Sotomayor searches. Since Sotomayor does not disclose or suggest a user defining a subsection delimiter, Sotomayor cannot disclose or suggest searching a document to find occurrences of items corresponding to a subsection delimiter defined by a user.

Additionally, the Office Action stipulates that Sotomayor fails to disclose “a document input device operative to provide an electronic version of a document, search for text and a document divider operative to divide the document into subsections.” The Office Action also stipulates that, “Sotomayor also fails to teach

scanning the document to generate scanned document data and performing recognition functions on the scanned document data to generate a recognized version of the document.” Accordingly, the Office Action appears to rely on Saito for disclosure of these elements with regard to **claims 4, 5, 8, 10-13, 18 and 20-28**.

However, it is respectfully submitted that there is no motivation in the art, other than that which might have been gleaned from the present application, to combine elements of Saito with elements of Sotomayor. Sotomayor allegedly discloses automatic index creation for a word processor which allows an author to select an index generation function for a document being processed (edited) and then, using a semantic analyzer program running on a computer, (b) automatically identifies significant key topics within the document, (c) generates and embeds index tokens into the text of the document (Abstract). Since Sotomayor asserts that the semantic analyzer of Sotomayor is capable of automatically identifying significant key topics within the document, there is no motivation to combine the user defined search template of Saito with the subject matter of Sotomayor.

With regard to scanning and optical character recognition, it is respectfully submitted that Sotomayor is concerned with automatic index creation for a word processor and does not disclose or suggest that the methods of Sotomayor should or could be applied to a scanned and character-recognized document. Additionally, Saito does not disclose or suggest that the search template of Saito could or should be used in generating an index. Accordingly, the only motivation to combine Sotomayor and Saito is that gleaned from the present application, and the rejection of **claims 4, 5, 8, 10-13, 18 and 20-28** are based on impermissible hindsight.

Furthermore, it is respectfully submitted that the Office has not met its burden for suggesting a motivation to combine the references. Conclusory statements, such as those provided in the Office Action, do not fulfill the agency's obligation. “The Board cannot rely on conclusory statements when dealing with particular combinations of prior art and specific claims, but must set forth the rationale on which it relies” (MPEP 2144.03, *In re Lee*, 277 F.3d 1338, 1344-45, 61 USPQd 1430, 1443-35 (Fed. Cir. 2002)).

For at least the foregoing reasons, **claims 4, 5, 8, 10-13, 18 and 20-28** are not anticipated and are not obvious in view of Sotomayor and Saito.

Additionally, **claim 8** recites displaying a plurality of document pages on a user interface, selecting at least one demarcation point on at least one of the plurality of pages and using the at least one demarcation point as the defined subsection delimiter. It is respectfully submitted that Saito does not disclose or suggest displaying a plurality of document pages on a user interface and selecting at least one demarcation point on at least one of the plurality of pages and using the at least one demarcation point as the defined subsection delimiter. Even if Saito discloses a user selecting a sub area, Saito does not disclose or suggest using a selected point as the defined subsection delimiter. Instead, Saito discloses a search template including not only coordinates but also an indentation parameter, font size parameter, font type and a number of lines. Accordingly, Saito does not disclose or suggest using the at least one demarcation point as the defined subsection delimiter.

**Claim 10** recites *inter alia*: a document divider operative to divide the document into subsections based on the recorded information regarding the occurrences corresponding to the delimiter definition. As pointed out by the Office Action, Saito discusses a division unit 205 that divides a document into separate sub areas. However, the division unit of Saito does not divide the document into sub areas based on recorded information regarding occurrences corresponding to a delimiter defined by a user. Indeed, the division unit of Saito divides the document into separate sub areas in order to present the user with sub units to select from, whereby such selection is used to generate the search template of Saito. For instance, column 4, lines 37-44, which indicates that the logic structure construction unit allows a user to select the divided areas where desired information is contained, column 5, line 54 - column 6, line 6, which indicates that a division unit 205 first divides the document image into separate sub areas, which respectively include characters, figures, data grams, tables and photographs. This occurs before the optical character recognition unit 207 converts alphanumeric characters in the character sub areas into computer recognizable data such as an ASCII format. All of this occurs without any discussion of user input. User defined input is not considered until element **802** of FIG. 5 (see column 6, lines 29-36). Accordingly, it is respectfully submitted that the combination of Sotomayor and Saito does not disclose or suggest at least the document divider recited in **claim 10**. Therefore,



**claim 10**, as well as **claims 11-13**, which depend therefrom, is not anticipated and is not obvious in light of Sotomayor and Saito.

Regarding **claim 11**, arguments similar to those submitted in support of **claim 2** are submitted in support of **claim 11**. Even if Sotomayor discloses an author including heading tokens in a document, Sotomayor does not disclose or suggest a system including a delimiter designator that is operative to receive an indication from the author or user that heading tokens or a predefined machine-readable symbol be used as a delimiter for searching a document for occurrences corresponding to the delimiter.

Furthermore, even if Saito indicates that a search template includes coordinates, an indentation parameter, a font size parameter, a font type and number of lines indication, Sotomayor does not disclose or suggest a user indicating any of that information or a delimiter designator that is operative to accept an indication of at least one of those or a font size, font style, text string, text location description, predefined machine-readable symbol and a specific point coordinate within the document as a delimiter designation. Instead, Saito only discloses a user selecting from among a plurality of sub-areas.

**Claim 20** recites *inter alia*: defining a subsection delimiter comprises marking a paper version of the document with at least one predetermined machine-readable demarcation symbol prior to scanning the document. It is respectfully submitted that Sotomayor and Saito do not disclose or suggest marking a paper version of the document with at least one predetermined machine readable demarcation symbol prior to scanning the document. Furthermore, the Office Action does not assert that Sotomayor and/or Saito disclose or suggest at least this element of **claim 20**.

**Claim 22** recites *inter alia*: determining a subsection delimiter definition including at least one delimiter characteristic wherein determining the subsection delimiter comprises a user indicating at least one of a font size and font style. It is respectfully submitted that the combination of Sotomayor and Saito does not disclose or suggest a user indicating at least one of a font size and a font style. Even if Saito discloses that a search template includes a font size and a font type, Saito does not disclose or suggest a user indicating one or both of these. Instead, the user of Saito selects a sub-area. Further in this regard, it is respectfully submitted that since Saito does not allow a user to designate appropriate

characteristics, the system of Saito is trained on a plurality of documents so that characteristic weights can be associated with the elements of a search template (e.g., FIG. 17, FIG. 18). Accordingly, determining a search template according to Saito is far more complicated and time consuming than defining a subsection delimiter according to the method recited, for example, in **claim 22**.

Arguments similar to those submitted in support of **claim 22** are submitted in support of **claim 23**. Additionally, it is respectfully submitted that the combination of Saito and Sotomayor does not disclose or suggest the user indicating a specific point coordinate within the document.

**Claims 24 and 25** depend from **claims 22 and 23**, respectively, and are not anticipated and are not obvious for at least those reasons.

**Claims 26 and 27** depend from **claim 10** and are not anticipated and are not obvious for at least that reason. Additionally, arguments similar to those submitted in support of **claims 22 and 23** are submitted in support of **claims 26 and 27**.

**Claim 28** recites *inter alia*: defining a subsection delimiter, wherein defining the subsection delimiter comprises at least one of a user building a subsection delimiter from a list of predetermined potential subsection delimiter components, performing statistical analysis on recognized characteristics to select characteristics that are most likely to be associated with subsection delimiters, a user entering a subsection delimiter by selecting symbols on a displayed portion of the electronic version of a document, and designating at least one demarcation point on at least one displayed portion of the electronic document to create a list of demarcation points to be used as a set of delimiter definitions. It is respectfully submitted that the Office Action does not assert that Sotomayor, Saito or a combination thereof discloses or suggests at least this portion of **claim 28**. Additionally, **claim 28** recites searching the recognized version to find occurrences of items that correspond to the defined subsection delimiter and using the found items to separate the document into separate sections. As explained above, the sub-areas of Saito are not found based on a subsection delimiter defined by a user.

For at least the foregoing additional reasons, **claims 8, 10-13, 18 and 20-28** are not anticipated and are not obvious in view of Sotomayor and Saito.

**Claim 6** was rejected under 35 U.S.C. §103(a) as being unpatentable over Sotomayor in view of Kuga. However, **claim 6** depends from **claim 2** and is patentably distinct and not obvious for at least that reason.

**Claims 15 and 16** were rejected under 35 U.S.C. §103(a) as being unpatentable over Sotomayor in view of Saito and further in view of Schmidt. **Claim 17** was rejected under 35 U.S.C. §103(a) as being unpatentable over Sotomayor, Saito, Schmidt and Herregods. However, **claim 15** depends from **claim 10**. **Claims 16 and 17** depend from **claim 15**. Therefore, **claims 15, 16 and 17** depend from **claim 10** and are not anticipated and are not obvious in view of Sotomayor, Saito, Schmidt and Herregods for at least that reason.

#### Telephone Interview

In the interests of advancing this application to issue the Applicant respectfully requests that the Examiner telephone the undersigned to discuss the foregoing or any suggestions that the Examiner may have to place the case in condition for allowance.

#### CONCLUSION

**Claims 2, 4-8, 10-13, 15-18 and 20-28** remain in the application. **Claims 2, 22, 23 and 28** have been amended.

For at least the foregoing reasons, the application is in condition for allowance. Accordingly, an early indication thereof is respectfully requested.

Respectfully submitted,

FAY, SHARPE, FAGAN,  
MINNICH & McKEE, LLP

10/17/06  
Date

Patrick R. Roche  
Patrick R. Roche, Reg. No. 29,580  
Thomas Tillander, Reg. No. 47,334  
1100 Superior Avenue, 7<sup>th</sup> Floor  
Cleveland, Ohio 44114-2579  
(216) 861-5582